

Figure A

cDNA sequence of SEQ ID NO: 2076 with the start and stop codons in bold and boxed, and the intron splice sites indicated with //. The three exons are underlined. The 5' and 3' untranslated regions are in italics.

*CAAACGTCTCCGTTTCTCTCCAAGCTGAAC***ATG**GACAAGAAGCCAGACGACGACAGTGGTAAGTCCCAAGATGTCGAGGTGA
GAAAAGGGCCGTGGACGATGGAAGAGGATCTCATCCTCATCAACTACATAGCGAATCACGGCGAAGGCAGTTGGAACCCCTA
GCCAAAGCTGCTGG//TCTAAACGTACCGGGAAGAGTTGTCTGGCTCCGGTGGCTGAACTATCTGCGACCCGACGTCCGGAGA
GGCAACATCACTACTGAGGAGCAGCTCCTGATCATGGAACTGCATGCCAAGTGGGGAAACAGGT//GGTCTAAAATTGCAAG
CATCTTCCCGGAAGGACTGACAATGAGATAAGACTTCTGGAGGACTAGAATCCAAAGCACATCAAGCAAGCAGAGGCTTT
CTCTGGTCAGAGCTCCGAGATGAGTGATCAAGCAAGCACAAGCCACATGTCCAGCATGCCAGAGCCGATGGAGACCTACGACT
CACCGCCGTCATTCCAAGGCAACAACAACATGGAGCCTTTGCCGGTGAATTTGTCTGGTCGAGTCAAAATGAAGCCTACTGGAGC
ATGGACGATCTTTGGTCTATGCAGTTACTCAATGGGGATTGATCGCGGGTGACGTATGGTGCAGTAATCGAAATGGTTCGTT
TACAATAATAGCTAGGTTTGTTTACATAAAATGGACATTAGCTTTTATCTCACATATATATCTACATACATGTGCTAGTTTA
GAAGTTATCTACAAATATGTGCATGAGTTGTAAACGAAACTACCATCTGCAGTTTGCATCCCCGCTATGTAATGACTGAAATA
ATCAAGCGAGATTATTTGGCTTAAAAAAAAAAAA

Figure B

CLUSTAL X (1.8) multiple sequence alignment of the amino acid translation of SEQ ID NO 2076 with the poplar gene prediction using Twinscan (Washington University) with a locally assembled Poplar contig from Phrap (University of Washington) alignment of genomic Poplar sequences (*Populus balsamifera*) made available by the Department of Energy Joint Genome Institute (University of California) and *A. thaliana* homologues (The Arabidopsis Information Resource (TAIR) Accession Nos At5g40350.1 and At3g01530.1).

```

Poplar      ----MDKS----P-----CNSQDVEVRKGPWTLKEDLILITNYIANHGEGVWNSLAKAAGLK
SEQ NO 2076 ----MDKK----PDDDSGKSQDVEVRKGPWTMEEDLILINYIANHGEGSWNSLAKAAGLK
At5g40350.1 ----MEKR----ESSGSGSGSDAEVRKGPWTMEEDLILINYIANHGEGVWNSLAKSAGLK
At3g01530.1 METTMKKKGRVKATITSQKEEGTVRKGPWTMEEDFILFNYILNHGEGLWNSVAKASGLK
          *.*          :  *****;***;** *** ***** **;*;***

Poplar      RTGKSCRLRWLNLYLRPDLRGNITPEEQLLIMELHAKLGNRWSKIAKHLPGRTDNEIKNY
SEQ NO 2076 RTGKSCRLRWLNLYLRPDVRRGNITTEEQLLIMELHAKWGNRWSKIAKHLPGRTDNEIKNF
At5g40350.1 RTGKSCRLRWLNLYLRPDVRRGNITPEEQLLIMELHAKWGNRWSKIAKHLPGRTDNEIKNF
At3g01530.1 RTGKSCRLRWLNLYLRPDVRRGNITTEEQLLIQLHAKLGNRWSKIAKHLPGRTDNEIKNF
          *****;***** **** *:**** *****;*****;

Poplar      WRTRIKKHTK---QTEFFAAGSSETNEHGSSTCQVS-SATDQMETY-CPPFYQG--DV-G
SEQ NO 2076 WRTRIQKHIK---QAEAFSGQSSEMSDQ-ASTSEMS-SMPEPMETYDSPPSFQGNMNM-E
At5g40350.1 WRTKIQKYIIKSGETTTVGSSQSEFINHHATTSEVMNDTQETMDMYSPTTSYQHASNINQ
At3g01530.1 WRTKIQRHMK--VSENMMNHQHHCSGN-SQSSGMT-TQSSGKAIDTAESFSQAKTT--
          ***:***:      :. . . : : :. . . :.

Poplar      AFSGGN-----IPQELNE-NYWSMEDLWSMQLLNGD----
SEQ NO 2076 PLP-VN-----LSVESNE-AYWSMDDLWSMQLLNGD----
At5g40350.1 QLNYGNYVPESGSIIMPLSVDQSEQNYWSVDDLWPMNIYNGN----
At3g01530.1 TFN-----VVEQQSNE-NYWNVEDLWPVHLLNGDHEVI
          :          :  :.* **.:***.:**

```

Figure C

Alignment of cDNAs of SEQ ID NO: 2076, *A. thaliana* At5g40350, Poplar spp.

Predicted intron/exon boundaries (underlined and bold), based on the *Arabidopsis thaliana* At5g40350 splicing sites and confirmed from a gene prediction using Twinscan with a locally assembled Poplar contig from Phrap alignment of genomic Poplar sequences (*Populus balsamifera*) made available by the Department of Energy Joint Genome Institute, are identified for poplar and SEQ ID NO: 2076. The splice sites are indicated by //.

CLUSTAL X (1.8) multiple sequence alignment

```

SEQ ID NO 2076      -----ATGGACAAGAAGCCAGACGACGACAGT
Poplar              -----ATGGATAAAAGTCCATGCA---AC---
At3g01530.1         ATGGAGACGACGATGAAGAAGAAAGGGAGAGTGAAAGCGACAATAACGTC
At5g40350.1         -----ATGGAGAAAAGAGAAAGTAGTGGTGG-
                      * * * * *

SEQ ID NO 2076      G-GTAAAGTCCCAAGATGTCGAGGTGAGAAAAGGGCCGTGGACGATGGAAG
Poplar              -----TCTCAGGATGTTGAAGTGAGAAAAGGGCCATGGACCTTGAAG
At3g01530.1         ACAGAAAGAAGAAGAAGGAACAGTGAGAAAAGGACCTTGGACTATGGAAG
At5g40350.1         GTCTGGATCAGGAGATGCAGAGGTGAGAAAAGGGCCATGGACGATGGAAG
                      * * * * *

SEQ ID NO 2076      AGGATCTCATCTCATCAACTACATAGCGAATCACGGCGAAGGCAGTTGG
Poplar              AAGACTTGATCTTAACCAACTACATCGCGAACCATGGTGAAGGTGTATGG
At3g01530.1         AAGATTTTCATCTCTTTAATTACATCCTTAATCATGGTGAAGGTCTTTGG
At5g40350.1         AAGATTTGATTCTCATCAATTATATCGCCAATCATGGTGAAGGTGTTTGG
                      * * * * *

SEQ ID NO 2076      AACTCCCTAGCCAAAGCTGCTG//GTCTAAACGTACCGGGAAGAGTTGTCC
Poplar              AACTCGCTTGCAAAAGCTGCGAG//GTCTGAAACGTACCGGGAAGAGTTGTAG
At3g01530.1         AACTCTGTGCGCCAAAGCCTCTG//GTCTAAACGTACTCGAAAAGTTGTCC
At5g40350.1         AACTCTCTCGCCAAATCTGCGAG//GACTAAACGCACCGGGAAGTTGCCG
                      *****

SEQ ID NO 2076      GCTCCGGTGGCTGAACTATCTGCGACCCGACGTCCGGAGAGGCAACATCA
Poplar              GCTCCGTTGGCTCAACTACTTGCGGCCTGACCTTCGAAGAGGGAATATTA
At3g01530.1         GCTCCGGTGGCTGAACTATCTCGACCCAGATGTGCGGCGAGGGAACATAA
At5g40350.1         GCTCCGGTGGCTGAACTACCTCCGACCTGATGTGCGACGGGGAATATCA
                      *****

SEQ ID NO 2076      CTACTGAGGAGCAGCTCCTGATCATGGAACCTGCATGCCAAGTGGGGAAC
Poplar              CTCCTGAAGAACAGCTCTTGATCATGGAACCTGCATGCTAAGTTGGGAAAC
At3g01530.1         CCGAAGAAGAACAGCTTTTGATCATTCAGCTTCATGCTAAGCTTGGAAAC
At5g40350.1         CACCAGAAGAACAGCTCACCATCATGGAACCTTCATGCAAAATGGGGAAT
                      * * * * *

SEQ ID NO 2076      AG//GTGGTCTAAAATTGCAAAGCATCTTCCCGGAAGGACTGACAATGAGAT
Poplar              AG//GTGGTCTGAAAATTGCAAAGCATCTTCCAGGGAGGACCGACAACGAGAT
At3g01530.1         AG//GTGGTCTGAAGATTGCGAAGCATCTTCCGGGAAGAACGGACAACGAGAT
At5g40350.1         AG//GTGGTCTGAAAATTGCAAAGCATTTACCAGGAAGAACCGACAATGAGAT

```

** ***** ** ***** ***** * * * * * ** ***** **

SEQ ID NO 2076
Poplar
At3g01530.1
At5g40350.1

AAAGAACTTCTGGAGGACTAGAAATCCAAAAGCACATCAAGCAAGCAGAGG
AAAGAACTTCTGGAGGACTAGAAATCAAGAAGCATACTAAGCAAACCTGAAC
AAAGAACTTCTGGAGGACAAAGATTTCAGAGACACATGAAGTGTTCATCGG
AAAGAACTTTTGGAGGACTAAGATCCAGAAATACATCATCAAGAGCCGAG
***** * ***** * * * * *

SEQ ID NO 2076
Poplar
At3g01530.1
At5g40350.1

CTTCTCTGGTC-----AGAGCTCCGAGATGAGTGATCAAGCAAG-
CATTTGCGGCGG-----GGAGTTCTGAGACTAATGAACATGGGAG-
AAAATATGAT-----GAATCATCAACATCATTGTTCCGGAAACT
AAACGACGACCGTTGGATCACAAGCTCCGAGTTTATAAACCATCATGC-
* * * *

SEQ ID NO 2076
Poplar
At3g01530.1
At5g40350.1

CACAA---GCCACATGTCCAGCATGC--CA-GAGCCGATGGAGACCTACG
CAGTACTTGCCAAGTTTCCAGCGCAA--CC-GACCAAATGGAGACCTATT
CACAGAGCTCGGGGATGACGACGCAAGGCA-GCTCCGGCAAAGCCATAGA
GACAACGAGCCATGTTCATGAATGATACTCAAGAAACCATGGATATGTATT
* * * *

SEQ ID NO 2076
Poplar
At3g01530.1
At5g40350.1

ACTCACCGCCG--TCATTCCAA-----GGCAACAACAACATGGAGCCTT
GTCCACCA-----TTCTATCAA-----GGAGACGTA-----GGGGCTTT
CACGGCTGAGAGCTTCTCTCAG-----GCGAAGACGA---CGACGTTTA
CTCCAACGACG--TCGTATCAACATGCCAGCAATATTAATCAGCAGCTTA
* * * * *

SEQ ID NO 2076
Poplar
At3g01530.1
At5g40350.1

-TGCCGGTG--AATTTGTCGGTCG-----
-TTCTGGTGGGAATATACCTCAAG-----
-ATGTGGTG--AA--CAAC-----
ATTATGGTAATTATGTGCCTGAATCCGGTTCGATCATGATGCCATTATCT
*** * *

SEQ ID NO 2076
Poplar
At3g01530.1
At5g40350.1

-----AGTCAAATGAAGCCTACTGGAGCATGGACGATCTTTGGTCTAT
-----AACTGAACGAAAATATTGGAGCATGGAGCATCTCTGGTCCAT
-----AGTCAAACGAGAATTACTGGAACGTTGAAGATCTGTGGCCCGT
GTTGATCAATCCGAACAAAATATTGGAGCGTCGATGATCTTTGGCCCAT
* * * * *

SEQ ID NO 2076
Poplar
At3g01530.1
At5g40350.1

GCAGTTACTCAATGGGGATTGA-----
GCAACTACTTAATGGCGAT-----
CCACTTGCTTAATGGTGACCACCATGTGATTAA
GAATATATATAATGGTAATTAA-----
* * ***** *

Figure D**Results from a BLASTX homology search using SEQ ID NO 2076:**

At5g40350.1 myb family transcription factor / similar to Myb26
 GI:1841475 from [Pisum sativum]; supported by
 full-length cDNA: Ceres: 262460.
 Length = 214

Score = 240 bits (606), Expect = 6e-64
 Identities = 125/209 (59%), Positives = 146/209 (69%), Gaps = 32/209 (15%)

Query: 8 SQDVEVRKGPWTL EEDLILITNYIANHGE GVMNSLAKAAGLKRTGKSCRLRWLN YLRPDLR 67
 S D EVRKGPWT+ EEDLIL NYIANHGE GVMNSLAK+AGLKRTGKSCRLRWLN YLRPD+R
 Sbjct: 12 SGDAEVRKGPWTL EEDLILIN YIANHGE GVMNSLAKSAGLKRTGKSCRLRWLN YLRPDVR 71

Query: 68 RGNITPBEQLLIMELHAKLGNRWSKIAKHLPGRTDNEIKNYWRTRIKCH---TKQTRPFA 124
 RGNITPBEQL IMELHAK GNRWSKIAKHLPGRTDNEIKN+WRT+I+K+ + +T
 Sbjct: 72 RGNITPBEQLTIMELHAKGNRWSKIAKHLPGRTDNEIKNFWRTKIQKYIISKGETTTVG 131

Query: 125 AGSSEFTNHRGSSSTQVSSATDQ-MET YCP-PFYQGDVGAPSGGNI PQELN----- 172
 + SSB H ++T V + T + M+ Y P YQ NI Q+LE
 Sbjct: 132 SQSSEFTNHHAATTSVMNDTQETIMNYSPITTSYQ-----HASNINQQLNYGNYVPESSGS 185

Query: 173 -----ENYWSMEDLNSMQLNGD 190
 +NYWS++DLW M + NG+
 Sbjct: 186 IMMPLSVDQSEQNYWSVDDLWPMNIYNGN 214

At3g01530.1 myb family transcription factor / contains PFAM
 profile: myb DNA binding domain PF00249 ;supported by
 full-length cDNA: Ceres:94595.
 Length = 206

Score = 233 bits (588), Expect = 8e-62
 Identities = 118/196 (60%), Positives = 139/196 (70%), Gaps = 13/196 (6%)

Query: 3 KSPCNSQDVE---VRKGPWTL EEDLILITNYIANHGE GVMNSLAKAAGLKRTGKSCRLRWL 59
 K+ SQ B VRKGPWT+ EED IL NYI NHGEG+MNS+AKA+GLKRTGKSCRLRWL
 Sbjct: 12 KATITSQKEEBGTVRKGPWTMEEDFILPNYILNHGEGLWNSVAKASGLKRTGKSCRLRWL 71

Query: 60 NYLRPDLRRGNITPBEQLLIMELHAKLGNRWSKIAKHLPGRTDNEIKNYWRTRIKCHTKQ 119
 NYLRPD+RRGNIT EEQLLI++LHAKLGNRWSKIAKHLPGRTDNEIKN+WRT+I++H K
 Sbjct: 72 NYLRPDVRRGNITPBEQLLIQLNAKLGNRWSKIAKHLPGRTDNEIKNFWRTKIQRHMKV 131

Query: 120 TEPFAAGSSSEFTNHRGSSSTQVSSATDQMET-----YCPFPFYQGDVGAPSGGNI PQELNEN 174
 + + H S Q S T Q + F Q F+ + Q+ NEN
 Sbjct: 132 S---SENMMNHQHCSCNSQSSGMITQSSGKAIDTAESFSAKTITFN--VVEQQSSEW 186

Query: 175 YWSMEDLNSMQLNGD 190
 YW++EDLW + LLNGD
 Sbjct: 187 YMWVEDLWVHLLNGD 202

Figure E

Phylogenetic comparison of SEQ ID NO: 2076 with *A. thaliana* homologues At3g01530.1 and At5g40350.1, and Poplar (*P. balsamifera*, gene prediction using Twinscan (Washington University) with a locally assembled Poplar contig from Phrap (University of Washington) alignment of genomic Poplar sequences made available by the Department of Energy Joint Genome Institute (University of California). This tree was generated using NJPlot.

